wwPDB EM Validation Summary Report (i)

Jul 3, 2024 – 12:16 pm BST

PDB ID	:	7PEL
EMDB ID	:	EMD-11052
Title	:	CryoEM structure of simian T-cell lymphotropic virus intasome in complex with PP2A regulatory subunit B56 gamma
Authors	:	Barski, M.; Pye, V.E.; Nans, A.; Cherepanov, P.; Maertens, G.N.
Deposited on	:	2021-08-10
Resolution	:	3.34 Å(reported)
Based on initial models	:	6TJU, 6QBW, 5JJA, 6TOQ

This is a wwPDB EM Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

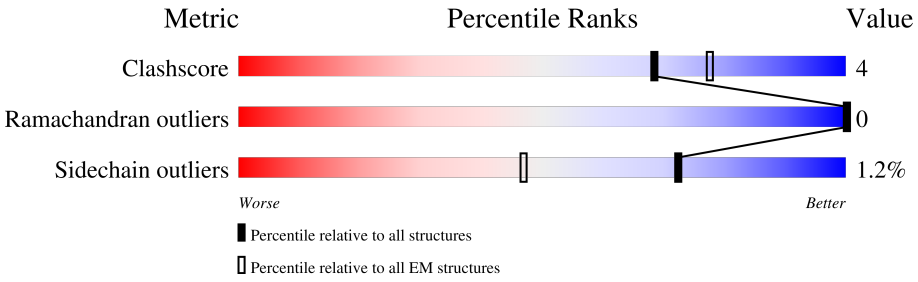
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.34 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



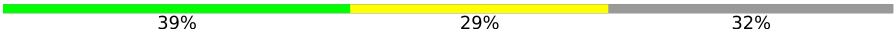
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	301	76% 9% • 14%
1	B	301	88% 5% 8%
1	D	301	75% 9% • 14%
1	E	301	87% 6% 8%
2	C	697	38% 5% 58%
2	F	697	38% • 58%
3	K	30	40% 30% 30%
3	M	30	40% 30% 30%
4	L	28	43% 25% 32%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	N	28	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 14998 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Pol protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	258	Total	C	N	O	S	0	0
			2053	1307	384	355	7		
1	B	278	Total	C	N	O	S	0	0
			2208	1399	417	384	8		
1	D	258	Total	C	N	O	S	0	0
			2053	1307	384	355	7		
1	E	278	Total	C	N	O	S	0	0
			2208	1399	417	384	8		

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP Q4QY51
A	-2	PRO	-	expression tag	UNP Q4QY51
A	-1	GLU	-	expression tag	UNP Q4QY51
A	0	PHE	-	expression tag	UNP Q4QY51
B	-3	GLY	-	expression tag	UNP Q4QY51
B	-2	PRO	-	expression tag	UNP Q4QY51
B	-1	GLU	-	expression tag	UNP Q4QY51
B	0	PHE	-	expression tag	UNP Q4QY51
D	-3	GLY	-	expression tag	UNP Q4QY51
D	-2	PRO	-	expression tag	UNP Q4QY51
D	-1	GLU	-	expression tag	UNP Q4QY51
D	0	PHE	-	expression tag	UNP Q4QY51
E	-3	GLY	-	expression tag	UNP Q4QY51
E	-2	PRO	-	expression tag	UNP Q4QY51
E	-1	GLU	-	expression tag	UNP Q4QY51
E	0	PHE	-	expression tag	UNP Q4QY51

- Molecule 2 is a protein called Isoform 3 of PC4 and SFRS1-interacting protein,Isoform Gamma-2 of Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	296	Total	C	N	O	S	0	0
			2423	1593	397	425	8		
2	F	296	Total	C	N	O	S	0	0
			2423	1593	397	425	8		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-316	SER	-	expression tag	UNP O75475
C	10	PHE	-	linker	UNP O75475
F	-316	SER	-	expression tag	UNP O75475
F	10	PHE	-	linker	UNP O75475

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*CP*TP*GP*TP*GP*TP*TP*TP*GP*GP*CP*GP*CP*TP*TP*CP*TP*CP*TP*C)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
3	K	21	Total	C	N	O	P	0	0
			422	204	66	132	20		
3	M	21	Total	C	N	O	P	0	0
			422	204	66	132	20		

- Molecule 4 is a DNA chain called DNA (5'-D(*GP*AP*GP*AP*GP*AP*AP*GP*CP*GP*CP*CP*AP*AP*AP*CP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
4	L	19	Total	C	N	O	P	0	0
			391	185	85	103	18		
4	N	19	Total	C	N	O	P	0	0
			391	185	85	103	18		

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

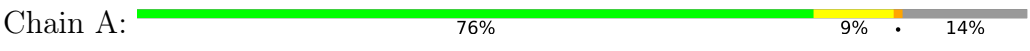
Mol	Chain	Residues	Atoms		AltConf
5	A	1	Total	Zn	0
			1	1	
5	B	1	Total	Zn	0
			1	1	
5	D	1	Total	Zn	0
			1	1	
5	E	1	Total	Zn	0
			1	1	

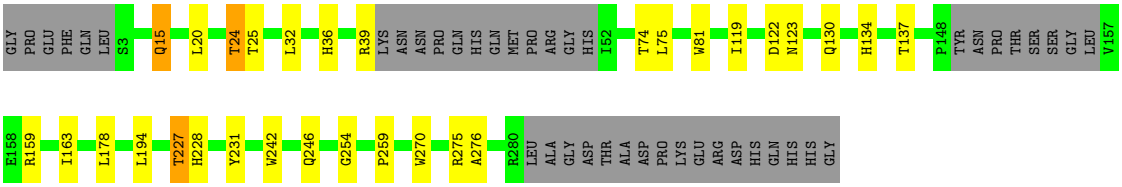
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Pol protein

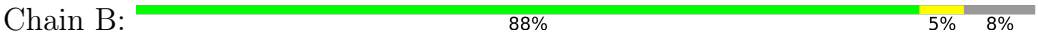
Chain A:

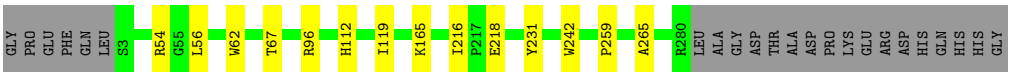




• Molecule 1: Pol protein

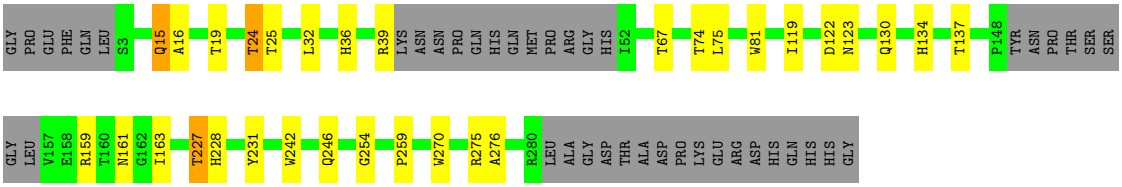
Chain B:





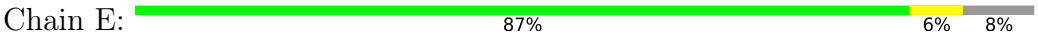
• Molecule 1: Pol protein

Chain D:



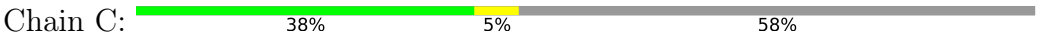
• Molecule 1: Pol protein

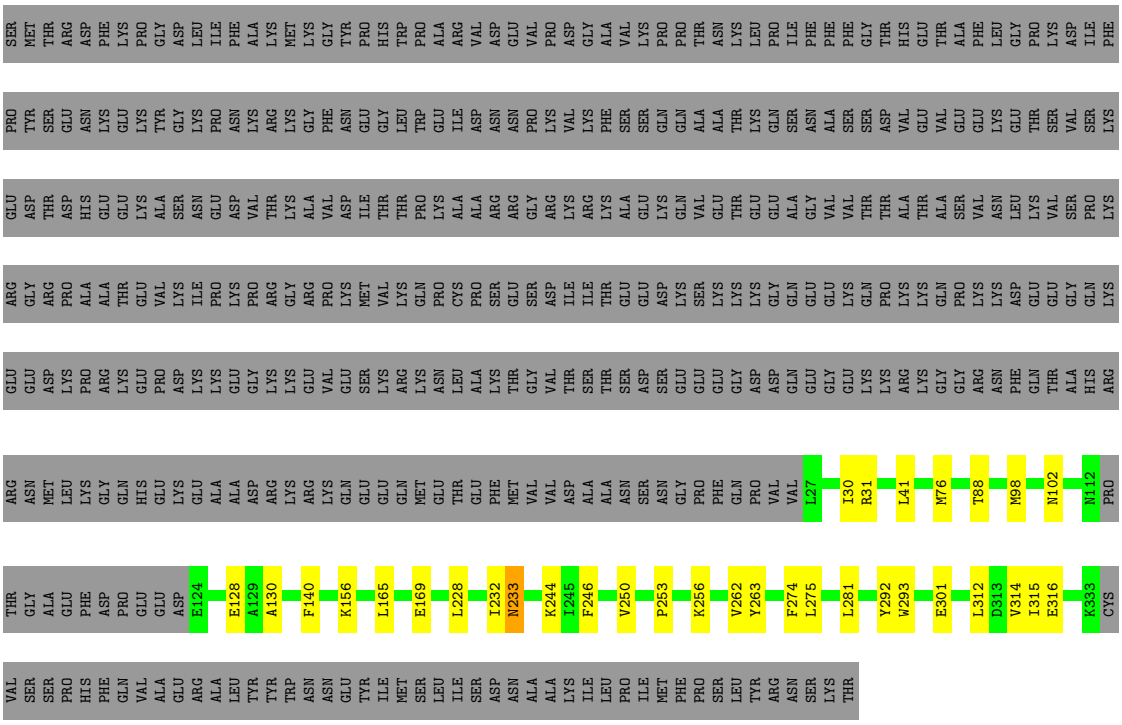
Chain E:



• Molecule 2: Isoform 3 of PC4 and SFRS1-interacting protein,Isoform Gamma-2 of Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform

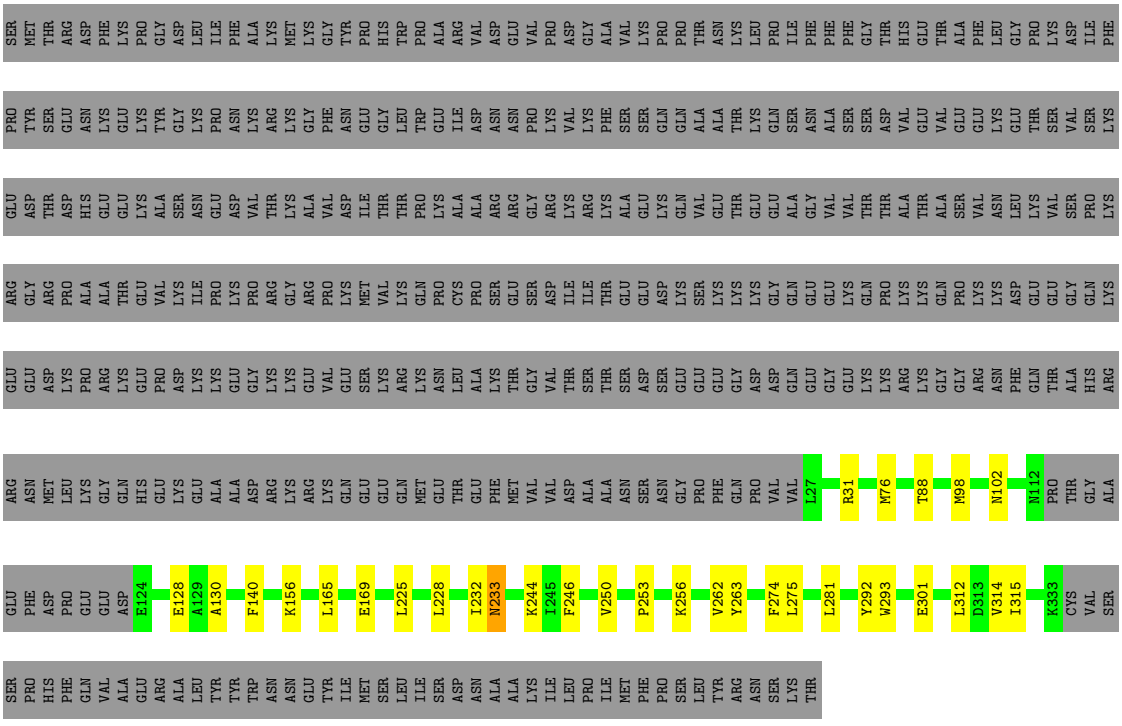
Chain C:





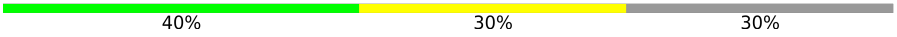
• Molecule 2: Isoform 3 of PC4 and SFRS1-interacting protein,Isoform Gamma-2 of Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform

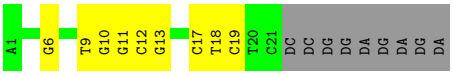
Chain F:



• Molecule 3: DNA (5'-D(*AP*CP*TP*GP*TP*GP*TP*TP*TP*GP*GP*CP*GP*CP*TP*TP*CP*TP*CP*TP*C)-3')

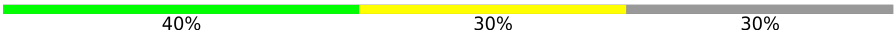
Chain K:

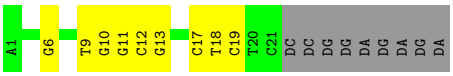




• Molecule 3: DNA (5'-D(*AP*CP*TP*GP*TP*GP*TP*TP*TP*GP*GP*CP*GP*CP*TP*TP *CP*TP*CP*TP*C)-3')

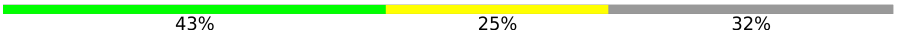
Chain M:





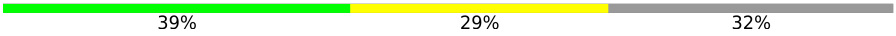
• Molecule 4: DNA (5'-D(*GP*AP*GP*AP*GP*AP*AP*GP*CP*GP*CP*CP*AP*AP*AP*CP *AP*CP*A)-3')

Chain L:



• Molecule 4: DNA (5'-D(*GP*AP*GP*AP*GP*AP*AP*GP*CP*GP*CP*CP*AP*AP*AP*CP *AP*CP*A)-3')

Chain N:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C2	Depositor
Number of particles used	161921	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50.4	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	FEI FALCON III (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.26	0/2116	0.43	0/2886
1	B	0.27	0/2279	0.44	0/3111
1	D	0.26	0/2116	0.43	0/2886
1	E	0.27	0/2279	0.44	0/3111
2	C	0.25	0/2488	0.40	0/3380
2	F	0.25	0/2488	0.40	0/3380
3	K	0.65	0/469	1.04	0/722
3	M	0.65	0/469	1.04	0/722
4	L	0.63	0/442	0.77	0/680
4	N	0.63	0/442	0.77	0/680
All	All	0.33	0/15588	0.52	0/21558

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2053	0	2040	17	0
1	B	2208	0	2180	8	0
1	D	2053	0	2040	16	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	2208	0	2180	10	0
2	C	2423	0	2422	20	0
2	F	2423	0	2422	18	0
3	K	422	0	242	6	0
3	M	422	0	242	7	0
4	L	391	0	211	6	0
4	N	391	0	211	7	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
5	D	1	0	0	0	0
5	E	1	0	0	0	0
All	All	14998	0	14190	111	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

The worst 5 of 111 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:244:LYS:HD3	2:C:281:LEU:HD13	1.63	0.81
2:F:244:LYS:HD3	2:F:281:LEU:HD13	1.63	0.81
1:E:218:GLU:OE1	2:F:31:ARG:NH1	2.26	0.69
1:B:218:GLU:OE1	2:C:31:ARG:NH1	2.26	0.68
1:B:54:ARG:NH2	3:K:6:DG:OP1	2.28	0.67

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	252/301 (84%)	240 (95%)	12 (5%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	276/301 (92%)	269 (98%)	7 (2%)	0	100	100
1	D	252/301 (84%)	240 (95%)	12 (5%)	0	100	100
1	E	276/301 (92%)	269 (98%)	7 (2%)	0	100	100
2	C	292/697 (42%)	277 (95%)	15 (5%)	0	100	100
2	F	292/697 (42%)	277 (95%)	15 (5%)	0	100	100
All	All	1640/2598 (63%)	1572 (96%)	68 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	222/258 (86%)	219 (99%)	3 (1%)	67	83
1	B	239/258 (93%)	237 (99%)	2 (1%)	81	90
1	D	222/258 (86%)	219 (99%)	3 (1%)	67	83
1	E	239/258 (93%)	237 (99%)	2 (1%)	81	90
2	C	265/628 (42%)	261 (98%)	4 (2%)	65	82
2	F	265/628 (42%)	261 (98%)	4 (2%)	65	82
All	All	1452/2288 (64%)	1434 (99%)	18 (1%)	72	84

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	88	THR
2	F	274	PHE
2	F	233	ASN
2	C	274	PHE
1	E	112	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.